Historic, Archive Document

Do not assume content reflects current scientific knowledge, policies, or practices.

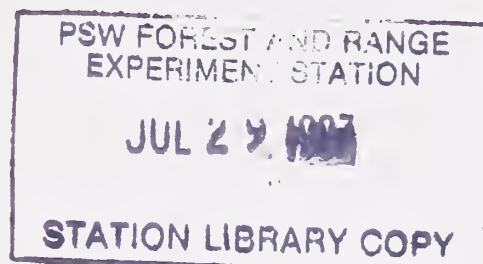Agriculture

Forest Service

**Rocky Mountain
Forest and Range
Experiment Station**

Fort Collins,
Colorado 80526

**Research Paper
RM-RP-331**

Robust Regression Analysis of Growth in Basal Area of Natural Pine Stands in Georgia and Alabama,

1962-72 and 1972-82

C. Y. Ueng

G. L. Gadbury

H. T. Schreuder

Abstract

Ueng, C.Y.; Gadbury, G.L.; Schreuder, H.T. **Robust regression analysis of growth in basal area of natural pine stands in Georgia and Alabama, 1962-72 and 1972-82.** Res. Pap. RM-RP-331. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station. 9 p.

Net growth and gross growth in basal area of selected plots of natural pine stands in Georgia and Alabama are examined under previously used models. Since large scale survey data may contain outlying observations, we use a procedure based on a linear model that is resistant to the influence of outliers. Our objective is to determine if the results of a previously used model hold when a linear model is fit to the data using our robust procedure. The data are drawn from forest inventory analysis measurements over two periods (cycle 4 and cycle 5). The analysis includes a bootstrap testing procedure. This testing procedure does not depend on parametric assumptions. Growth of the three species studied in Georgia consistently showed a significant decline from the first period to the second period. A similar but less consistent decline in growth was observed in Alabama.

Key words: forest inventory analysis, natural pine stands, bootstrap, robust procedure, growth decline

The Authors

C.Y. Ueng and **G.L. Gadbury** are Ph.D. students of statistics at Colorado State University in Fort Collins.

H.T. Schreuder is a mathematical statistician at the Rocky Mountain Forest and Range Experiment Station in Fort Collins.

Publisher

Cover photo: *Forty-year-old stand of loblolly pine. Photo courtesy of USDA Forest Service, Southern Research Station.*

Robust Regression Analysis of Growth in Basal Area of Natural Pine Stands in Georgia and Alabama, *1962-72 and 1972-82*

C.Y. Ueng

G.L. Gadbury

H.T. Schreuder

Contents

Robust Regression Analysis of Growth in Basal Area of Natural Pine Stands in Georgia and Alabama,

1962-72 and 1972-82

Introduction

Since 1928, the Forest Inventory and Analysis (FIA) units of the USDA Forest Service have surveyed the forest resources of the USA. Estimates of aggregates, such as areas in major land classes and/or forest types, volume by tree species, and changes in areas and volumes over time are obtained from the data. The surveys enable assessment of forest resource change over large areas. In recent decades, these inventories have produced data that, upon subsequent analyses, have revealed a decrease in growth of natural southern pine stands in the southeastern USA for the period 1972-1982 relative to 1962-1972 (Sheffield *et al*. 1985, Sheffield and Cost 1987). This reported decrease has been as high as 23% for loblolly pine (*Pinus taida L*.), 32% for Georgia shortleaf pine *(P. echinata Mill*.), 28% for Georgia slash *(P. elliotti)*, and 29% for Alabama longleaf *(P. palustris)* (Bechtold *et al*. 1991 and Ruark *et al*. 1991, hereafter Bechtold *et al*. and Ruark *et al*.).

These findings may cause concern because questions about potential causes for the reported decline in growth become relevant, and such data do not lend themselves to establishing cause-effect (Schreuder and Thomas 1991). For example, newspapers have discussed the possibility of a pollution effect even though such a link has certainly not been established.

Considering the seriousness of the implications of a growth decline reaching as high as 32%, a reanalysis of the two data sets used by Bechtold *et al*. and Ruark *et al*., respectively, is not only prudent but arguably necessary. Alternative analyses of a given data set are not unreasonable because statistical models are only approximations of reality, at best, and no single analysis can take into consideration all the different peculiarities associated with any given data set. The models used by Bechtold *et al*. and Ruark *et al*. have been subjected to careful scrutiny by others, but some issues remain unexplored. It is not unreasonable to expect some outlying observations in large scale survey data. If such observations are present in the above mentioned data sets, then to what extent do these outliers influence the results of the analyses? Would a more robust linear model based analysis reach conclusions that are consistent with prior findings? In addition, net growth (gross growth minus mortality) of natural pine stands has not been analyzed by other authors, and some question whether or not the mortality covariate is valid in a gross growth analysis. Therefore, we analyze gross growth both with and without the mortality term. So additional questions are, does an analysis of net growth still reveal growth declines from cycle 4 to cycle 5 and are any prior results affected by removing the mortality term from a gross growth analysis. All these questions have led us to reconsider the Bechtold *et al*. and the Ruark *et al*. data sets (hereafter referred to as data set #1 and data set #2, respectively).

We use a robust technique of fitting a linear model to data sets #1 and #2. It is an iterative procedure that fits a model, identifies outliers to the model, and modifies the outlying observations to make them more consistent with normal linear model theory. Then a new model is fitted to the modified data. The process continues until model coefficients change very little between iterations. So we would expect this procedure to be effective in dealing with the problem of outlier masking, where one or more outliers to a fitted model masks the presence of other outliers. To test for significance of a drop in tree growth, we use a nonparametric bootstrapping technique in addition to a classical procedure. The bootstrapping technique allows us an alternative way to check for significance in model coefficients without relying on normal linear model theory. Finally, we analyze both gross growth of trees and net growth (gross growth for data sets #1 and #2, and net growth for data set #1). For a gross growth analysis, we use the model form used by Bechtold *et al*. and Ruark *et al*. For net growth, however, we need to establish a proper form of a model. So we make use of generalized linear models (GLM's) to assist in determining the proper form. This class of models only require that the data follow an exponential family distribution rather than the normal distribution as in standard linear regression. We believe the procedures discussed in this paper should be valid in any situation where a large data set with suspected outliers is fitted to a linear model in order to detect the significance of one or more factors.

Review of Literature

Bechtold *et al.* (1991) classified a screened set of natural pine inventory plots in Georgia as loblolly, shortleaf, or slash pine (*P. elliottii* Engel.) and for each plot type they analyzed the data for "pines only" and "all trees." Thus, they analyzed a total of six data sets. Loblolly and slash pine were analyzed using the model:

$$\ln(G_i) = b_0{}^*c_1 + b_1{}^*c_2 + b_2{}^*S + b_3{}^*\ln(A) + b_4{}^*\ln(N) + b_5{}^*P + b_6{}^*\ln(M+1) + e_i,$$

$$i = 1, 2 \tag{1}$$

where

G_1 and G_2 are the gross growth rate of pines only and all trees respectively with diameters ≥ 1";

c_1, c_2 are "cycle" indicator variables of the survey periods 1962-1972 and 1972-1982 respectively;

A is the age of the stands;

N is the number of pines with dbh (diameter at breast height) ≥ 1";

S is site class; P is the ratio of pine basal area to total basal area;

M is the pine basal area mortality; b_j (j=1, ..., 6) are regression coefficients; and

e_i (i=1, 2) are error terms with mean zero and variance $\sigma_i{}^2$.

Shortleaf pine was analyzed with a model similar to model 1 but with an interaction between cycle and N rather than a common N term across cycles.

The basic assumption underlying model 1 is that since effects of the covariates on the growth rate do not change with the growth period, any difference in the growth rate not due to these covariates can be attributed to b_0 and b_1. Bechtold *et al.* tested the hypothesis that the gross growth rates of these two periods are equal after the effects of all other independent variables were eliminated, i.e., $H_0: b_0 = b_1$. For every analysis that Bechtold *et al.* completed, the null hypothesis was rejected showing a statistically significant drop in tree growth rates from cycle 4 to cycle 5.

Since the data analyzed by Ruark *et al.*, representing plots in both Georgia and Alabama, measured somewhat different variables in the two states, a slightly different model was used for their analyses:

$$\ln(G_i) = b_0{}^*c_1 + b_1{}^*c_2 + b_2{}^*S + b_3{}^*\ln(QMD) + b_4{}^*\ln(N) + b_5{}^*P + b_6{}^*\ln(M+1) + e_i,$$

$$i = 1, 2 \tag{2}$$

where G_1 and G_2 are the basal area growth of pine trees ≥ 1" (PSG1) and ≥ 5" (PSG5) respectively; QMD is the initial quadratic mean diameter of pine trees ≥ 1"; and other

variables are as defined for model 1. Ruark *et al.* fitted the model to data for loblolly pine, longleaf pine (*P. palustris* Mill), shortleaf pine, and slash pine (the last species only for Georgia). They compared the estimated growth rates between 1972-1982 with those estimated during the previous 10-year period and found reductions ranging from 3-31% in both states. All the results were statistically significant except for the 3% decline in natural loblolly pine in Alabama.

Ouyang *et al.* (1992) suggested improving test procedures applied by Bechtold *et al.* (1991) and Ruark *et al.* (1991) by using the same models but basing the tests on bootstrap and weighted jackknife confidence intervals. With the revised procedures, they confirmed the growth decline found by Bechtold *et al.* and Ruark *et al.* in Georgia and Alabama.

Van Deusen (1992), using increment core data collected from natural loblolly pine stands in the Piedmont regions of Georgia and South Carolina, fitted the following model to predict basal area growth:

$$BAG5 = b_0 + b_1{}^*DBH + b_2/AGE + b_3{}^*PBA + b_4{}^*NPBA$$

$$+ b_6{}^*PBA/PNT + e \tag{3}$$

where

BAG5 is five-year basal area growth in square feet centered around the survey measurement year;

DBH is diameter at breast height to the nearest 0.1 inch;

AGE is total age plus 3 computed from tree cores;

PBA is pine basal area to the nearest square foot;

NPBA is non-pine basal area;

PNT is the number of pine trees per acre; and

e is a random error term.

Van Deusen's results confirmed Bechtold et al.'s (1991) findings of a growth decline and also showed an unreported growth increase prior to the decline.

The Data

We use the same screened data sets used by Bechtold *et al.* and Ruark *et al.* (data sets #1 and #2). Bechtold *et al.* screened the Georgia data sets at the two measurement periods (1961-1972 and 1972-1982) to obtain similar stands of timber. The sample screening criteria are as follows:

1. Classified as timberland both at the beginning and end of the period.

2. All sample points located in the same forest condition.

3. No evidence of planting or artificial seeding.

4. No visual evidence of serious disturbance or treatment.

5. Classified as either a loblolly pine or shortleaf pine forest type in the Piedmont or Mountain Survey Units, or slash pine type in the Coastal Plain Units. To qualify as such, at least 50% of the initial basal area of trees > 2.5 cm (1.0 inches) had to be pine with the majority of the pine stocking in the cover-type of the species. Only 27 plots of loblolly, 14 of shortleaf, and 11 of slash pine were common to both periods since disturbance rates are high in these southern stands.

Data set #1 includes growth data for loblolly pine, shortleaf pine, and slash pine, and for each data set there is information for the growth of "all trees" and "pines only." We analyze both net growth and gross growth for data set #1.

Data set #2 covered the same data and periods as in data set #1 and the periods of 1962-1972 and 1972-1982 in Alabama. Ruark *et al.* used screening criteria that included items #3 and #5 (above) plus:

1. Sample points established at exactly the same location at both inventories.

2. No evidence of harvesting during the measurement period.

Again, few plots met the criteria for both periods.

Data set #2 includes growth data for loblolly pine, shortleaf pine, longleaf pine in both Alabama and Georgia with the latter also having growth data for slash pine. There is information for PSG1 growth and PSG5 growth (see the Review of Literature for a description of these). We analyze only gross growth for data set #2.

Modeling Gross Growth

For modeling the gross growth response, we decided to concentrate on models similar to those used by Bechtold *et al.* and Ruark *et al.* Van Deusen's (1992) tree model is inappropriate for our plot-level data. Gross growth is modeled with and without mortality as a covariate. So, for analyzing gross growth for data set #1, we use a model of the form of equation 1 (which we will refer to as model 1), and a model of the same form but without the term for mortality (which we refer to as model 1a).

For analyzing gross growth for data set #2, we use a model of the form of equation 2 (which we will refer to as model 2), and a model of the same form but without the term for mortality (which we refer to as model 2a). The specific robust fitting procedure will be presented in a later section.

Modeling Net Growth and the Generalized Linear Model Procedure

Overview of Generalized Linear Models

The response variable used in the analyses of Bechtold *et al.* (1991) and Ruark *et al.* (1991) is gross growth. However, for our analyses of net growth (gross growth - mortality), we are using a different response variable and, consequently, may need a model of a different form. Nelder and Wedderburn (1972) developed the theory and use of generalized linear models (GLMs), a general and flexible approach to model fitting. A GLM model is specified in its simplest form by the following:

(a) Independent observations $y_1, \ldots, y_n$ distributed according to an exponential family of distributions;

(b) Set of independent variables X (covariates) for each observation; and

(c) The link function g which relates the mean of an observation to the covariates (i.e., $g(\mu) = X^T\beta$).

We assume that (a) and (b) are met and concern ourselves with determining $g(\mu)$ in item (c). We use Pregibon's (1980) procedures to determine $g(\mu)$ under the assumption that requirements (a) and (b) are satisfied.

A generalized linear model procedure attempts to fit a function of the response variable to a linear model such as $g(Y) = X^T\beta + \varepsilon$ where g is the link function mentioned in (c), X is the matrix of covariates, and ε is a error vector. We start by assuming that g is contained within a family of power transformations, that is, g is of the form:

$$g(\mu; \alpha, \delta) = \frac{(\mu + \alpha)^\delta - 1}{\delta} \qquad (4)$$

and we assume an initial g, $g_0(\mu) = g(\mu; \alpha_0, \delta_0)$, where the true but unknown link function is denoted $g*(\mu) = g(\mu; \alpha*, \delta*)$. We choose our initial $g_0(\mu)$ to be the identity, that is $g_0(\mu) = \mu$, or $(\alpha_0, \delta_0) = (1,1)$. If we expand $g*(\mu)$ about our hypothesized link function $g_0(\mu)$ using a first-order Taylor series expansion, we get a form of $g*(\mu)$ that is:

$$g*(\mu) = g_0(\mu) + (\alpha*-\alpha_0) D_\alpha(g_0) + (\delta*-\delta_0) D_\delta(g_0) \qquad (5)$$

where $D_\alpha(g_0)$ and $D_\delta(g_0)$ are the first partial derivatives of $g(\mu; \alpha, \delta)$ in equation 4, with respect to α and δ respectively, that are evaluated at g_0. Hence, we may approximate the correct link $g*(\mu) = X^T\beta$ by $g_0(\mu) = X^T\beta + z^T\gamma$, where $z^T = (D_\alpha(g_0), D_\delta(g_0))$ and $\gamma^T = (\alpha*- \alpha_0, \delta*- \delta_0)$. It is

evident now that the difficulty of identifying g* is transformed into a problem of comparing two nested linear models to determine the best fit. It remains, however, to determine the four steps to estimate z, γ, and β. The steps are as follows:

1. Start with the assumed g_0 and fit the initial model $g_0(\mu) = X^T\beta$ to get estimates $\hat{\beta}$ and fitted values $g_0(\hat{\mu}) = X^T\hat{\beta}$ and $\hat{z} = z_{\beta=\hat{\beta}}$.

2. Fit a model using $\hat{z}$ as an additional set of explanatory variables (i.e. fit the model $g_0(\mu) = X^T\beta + \hat{z}^T\gamma$).

3. Compare the two deviances from the models in steps 1 and 2. If the difference is significant, the hypothesized link $g_0(\mu)$ in step 1 is not a good estimate. One can derive estimates of γ, α_* and δ_* from step 2, therefore obtaining a new link function.

4. Use this new link function estimate from step 3 and repeat steps 1-3 until a suitable estimate of the link function is obtained (i.e. the reduction in the deviance between steps 1 and 2 is not significant).

The Net Growth Model

We use the generalized linear model on 3 different forms of a net growth (NG) model on each data set from data set #1. Initially, we assume that g_0 is the identity function (i.e., $g_0(\mu) = \mu$), so our response variable is the untransformed net growth (NG). Other variables are as defined for model 1. The 3 forms of net growth robust models are

$$NG = b_0{}^*c_1 + b_1{}^*c_2 + b_2{}^*A + b_3{}^*\ln(N) + b_4{}^*S + b_5{}^*P \qquad (5)$$

$$NG = b_0{}^*c_1 + b_1{}^*c_2 + b_2{}^*\ln(A) + b_3{}^*\ln(N) + b_4{}^*S + b_5{}^*P \qquad (6)$$

$$NG = b_0{}^*c_1 + b_1{}^*c_2 + b_2{}^*\ln(A) + b_3{}^*\ln(N) + b_4{}^*S + b_5{}^*\exp(P) \qquad (7)$$

Results from this procedure, when used on these net growth models, determined that $g(\mu)$ is the identity link function for all data sets with the exception of loblolly "pine only" analysis which specified a log-link with an added constant to prevent taking logarithms of negative numbers. The form of the "best" net growth model was chosen from models 5 - 7 based on a quantity of the form of Akaike's Information criteria (AIC, Hastie and Tibshirani 1990) (table 1).

The Robust Fitting Procedure

Once the form of a linear model is obtained, we fit it to a data set using a robust procedure. It is well known that outlying observations can have an undue influence on the calculated model coefficients. Unfortunately, if observations are outliers and, according to the fit of one model, are removed from the data (or down weighted to reduce their influence), then fitting a new model to this altered data may produce a model with different coefficients. Then, according to this new model, additional observations that are outliers may occur that were not outliers in the first model. In short, this process could continue through several iterations. At each iteration, for example, a model is fit, outlying observations identified and modified in order to reduce their influence, and a new model fit to this altered data set. Hopefully, this process would eventually con-

Table 1. Details of the "best" net growth robust model for data set #1. Numbers in parantheses are the "best" models. Note that, for loblolly pines only, the response is in logged units.

	Slash/ pines only (7)	Slash/ all trees (6)	Shortleaf/ pines only (6)	Shortleaf/ all trees (5)	Loblolly/ pines only (7)[*]	Loblolly/ all trees (6)
c_1 (1st period)	1.4601	2.3254	0.4547	0.6863	1.4570	3.2276
c_2 (2nd period)	0.7810	1.5383	-0.4140	-0.5571	1.3588	2.3408
A				-0.0743		
ln(A)	-1.0306	-1.0034	-1.1426		-0.1991	-1.5845
ln(N)	0.4788	0.6213	0.5496	0.4329	0.0726	0.6085
S	0.0107	0.0048	0.0206	0.0234	0.0002	-0.0005
P		0.6494	1.5206	0.5246		2.2776
exp(P)	0.5630				0.2394	
R^2	0.2852	0.2112	0.3747	0.3113	0.3140	0.2529
$\hat{\sigma}$	1.6810	1.9660	1.6270	1.8450	0.3041	2.2460
AIC	6.8112	6.9819	6.4257	6.7325	4.1531	7.7032

[*] This model used ln(NG +3.4) for the response as determined by the GLM procedure.

verge (i.e., the same model is obtained at each iteration). This convergence occurred in the procedure we used to analyze data sets #1 and #2. The procedure involves an iterative process comprised of four steps (Chiu 1995):

Step 1: Set $Y^{(1)} = Y$ (the original response variable) and denote "k" as a counter of iterations. Initially, $k = 1$.

Step 2: Regress $Y^{(k)}$ on $\mathbf{X}$ (the covariates) to get predicted values $\hat{y}^{(k)}$, residual $\hat{\varepsilon}^{(k)}$, and standard deviation $\hat{\sigma}^{(k)}$.

Step 3: Compute the modified $Y^{(k+1)}$, such that

$$Y^{(k+1)} = \begin{cases} \hat{y}_i^{(k)} - 2^*\hat{\sigma}^{(k)} & \text{if } \hat{\varepsilon}_i^{(k)} < -2^*\hat{\sigma}^{(k)} \\ Y_i^{(1)} & \text{if } |\hat{\varepsilon}_i^{(k)}| \le 2^*\hat{\sigma}^{(k)} \\ \hat{y}_i^{(k)} + 2^*\hat{\sigma}^{(k)} & \text{if } \hat{\varepsilon}_i^{(k)} > 2^*\hat{\sigma}^{(k)} \end{cases}$$

Step 4: Let $k = k+1$ and repeat steps 2 and 3 until the model parameter estimates remain unchanged.

In step 3, the response variable is adjusted for outliers. Observations with estimated residuals larger than twice the square root of mean squared error (MSE) are adjusted to a value that is a distance of two times the square root of MSE away from the fitted value at that observation. Imagine a heavy tailed distribution truncated at a certain distance from the mean; conceptually, that is what is being done here. This robust fitting procedure is used to fit every model discussed in the prior sections.

The Testing Procedure

For all models we tested H_0: $(b_1-b_0) = 0$ (i.e., there is no difference in growth rates between the two periods) by using both a bootstrap method and a classical t-test. For the bootstrap test, 1000 bootstrap samples were drawn, and for each sample, the period effect between the two periods (b_1-b_0) was computed by fitting the chosen form of the model with the robust procedure. Confidence intervals for bootstrap samples were constructed based on the quantile of these 1000 sample period effects. A classical p-value was also calculated based on a t-test of H_0: $(b_1-b_0) = 0$.

For every model, predicted growth for each of the two periods was obtained by substituting the medians of the covariates in the fitted model. The percentage of the reduction was calculated based on this predicted growth.

Results and Discussion

Net Growth for Data Set #1

For the "best" (based on the AIC criterion) net growth model from models 5-7, we tested the hypothesis that the true link function was the identity function (i.e. $g_*(\mu) = \mu$). For all tree types in data set #1, the results suggest using the identity function for g for all tree types except loblolly pine where we chose to use $g(NG) = \ln(NG+3.4)$. Table 1 shows details of the "best" net growth robust models. Both the bootstrap and classical tests show significant growth differences for all six tree analyses ($\alpha = 0.05$). However, the bootstrap procedure did not detect a significant drop at $\alpha = 0.01$ for slash pine only and shortleaf pine only and the classical test results were not significant at $\alpha = 0.01$ for slash pine only and slash all trees (table 2). Net growth during the 2nd period was 22% less in slash-pines only, 14% less in slash-all trees, 45% less in shortleaf-pines only, 39% less in shortleaf-all trees, 43% less in loblolly-pines only, and 38% less in loblolly-all trees (table 3).

Table 2. Confidence intervals of (b_1-b_0) for bootstrap samples (1,000 samples) and classical p-values for testing H_0: $(b_1-b_0) = 0$ for the "best" net growth robust model for data set #1. Note that for loblolly pine, response is in logged units.

| Species/tree type | Confidence interval for bootstrap | | Classical p-value |
	99%	95%	
slash/pines only	(-2.4681, 0.1933)	(-2.1341,-0.4783)	0.0164
slash/all trees	(-2.3299,-0.1053)	(-1.9742,-0.2815)	0.0160
shortleaf/pines only	(-2.0221, 0.1829)	(-1.7639,-0.1005)	0.0051
shortleaf/all trees	(-4.3250,-1.6006)	(-4.0432,-2.0368)	0.0004
loblolly/pines only	(-0.2050,-0.0038)	(-0.1713,-0.0487)	0.0079
loblolly/all trees	(-1.8770,-0.0364)	(-1.6875,-0.2780)	0.0012

Table 3. *Estimated net growth from the "best" net growth robust models in table 1 for data set #1. All results are in original units.*

| Species/tree type | Estimated net growth | | |
	1st period	2nd period	Decrease (%)
slash-pines only	2.9456	2.3037	21.79
slash-all trees	3.6811	3.1674	13.96
shortleaf-pines only	3.1045	1.7130	44.82
shortleaf-all trees	4.1730	2.5641	38.55
loblolly-pines only	3.7155	2.1034	43.39
loblolly-all trees	5.0135	3.1277	37.61

Gross Growth for Data Set #1

Gross growth was analyzed using models 1 and 1a. We chose to focus on test results and omitted the model details obtained from the robust procedure. The 99% and 95% bootstrap confidence intervals and classical p-value for testing H_0: $b_1 - b_0 = 0$ for gross growth models (1) and (1a) are reported in tables 4 and 5 respectively.

All test results are significant at $\alpha = .05$ for models 1 and 1a. But at $\alpha = .01$, the bootstrap test is not significant in model 1 for slash pine only and shortleaf pine only, and the classical test is not significant for either of the two slash pine analyses in models 1 or 1a.

Table 4. *Confidence intervals of $(b_1 - b_0)$ for bootstrap samples (1000 samples) and classical p-values for testing H_0: $(b_1 - b_0) = 0$ for the gross growth robust model 1 for data set #1. For all results, the response is in logged units.*

| Species/tree type | Confidence interval for bootstrap | | Classical p-value |
	99%	95%	
slash/pines only	(-0.5588, 0.0089)	(-0.5114, -0.0685)	0.0258
slash/all trees	(-0.5522, -0.0205)	(-0.4758, -0.0445)	0.0259
shortleaf/pines only	(-0.5271, 0.0092)	(-0.4681, -0.0604)	0.0039
shortleaf/all trees	(-0.5105, -0.0329)	(-0.4543, -0.0975)	0.0009
loblolly/pines only	(-0.3403, -0.0786)	(-0.3155, -0.1096)	0.0000
loblolly/all trees	(-0.2985, -0.0385)	(-0.2594, -0.0707)	0.0000

Table 5. *Confidence intervals of $(b_1 - b_0)$ for bootstrap samples (1000 samples) and classical p-values for testing H_0: $(b_1 - b_0) = 0$ for the gross growth robust model 1a for data set #1. For all results, the response is in logged units.*

| Species/tree type | Confidence interval for bootstrap | | Classical p-value |
	99%	95%	
slash/pines only	(-0.5980, -0.0177)	(-0.5452, -0.0817)	0.0144
slash/all trees	(-0.5592, -0.0108)	(-0.4842, -0.0579)	0.0211
shortleaf/pines only	(-0.6292, -0.0324)	(-0.5237, -0.1020)	0.0022
shortleaf/all trees	(-0.5821, -0.0377)	(-0.5024, -0.1160)	0.0002
loblolly/pines only	(-0.3278, -0.0391)	(-0.3022, -0.0826)	0.0000
loblolly/all trees	(-0.2932, -0.0306)	(-0.2622, -0.0639)	0.0000

Table 6. *Estimated gross growth from the gross growth robust model 1 for data set #1. All results are in original units.*

| Species/tree type | Estimated net growth | | |
	1st period	2nd period	Decrease (%)
slash/pines only	3.5987	3.2635	9.31
slash/all trees	3.7024	3.3763	8.80
shortleaf/pines only	4.1997	2.6536	36.81
shortleaf/all trees	4.3321	3.4814	19.64
loblolly/pines only	5.3637	3.4208	36.22
loblolly/all trees	5.8039	4.0279	30.60

Table 7. *Estimated gross growth from the gross growth robust model 1a or data set #1. All results are in original units.*

| Species/tree type | Estimated net growth | | |
	1st period	2nd period	Decrease (%)
slash/pines only	3.5620	3.1884	10.49
slash/all trees	3.5806	3.2077	10.41
shortleaf/pines only	3.9195	2.6028	33.59
shortleaf/all trees	3.9864	3.2201	19.22
loblolly/pines only	5.1893	3.2915	36.57
loblolly/all trees	5.4054	3.6891	31.75

Tables 6 and 7 show the estimated gross growth for the 1st and 2nd periods and the percentage of the reduction from the 1st to 2nd period for gross growth robust models 1 and 1a respectively. Reductions in growth range from 9% for model 1 slash pine only to 37% for model 1 shortleaf pine only. All results from models 1 and 1a were similar.

Gross Growth for Data Set #2

Tables 8 and 9 show the 99% and 95% bootstrap confidence interval and classical p-value for testing $H_0: b_1 - b_0 = 0$

for gross growth models 2 and 2a respectively. Tables 10 and 11 show the estimated gross growth for the 1st and 2nd periods and the percentage of the reduction from the 1st to 2nd period for gross growth robust models 2 and 2a respectively.

The bootstrap and classical test results both show significant growth differences in the 14 state/species/tree type combinations at $\alpha = 0.05$ for the robust procedure for gross growth models 2 and 2a except for Alabama-loblolly-PSG5, Georgia-longleaf-PSG1, and Georgia-longleaf-PSG5 (tables 9 and 10). At $\alpha = 0.01$, the bootstrap test and classical result showed significant differences except for Alabama-shortleaf-PSG5, Alabama-loblolly-PSG1, Ala-

Table 8. Confidence intervals of $(b_1 - b_0)$ for bootstrap samples (1000 samples) and classical p-values for testing $H_0: (b_1 - b_0) = 0$ for the gross growth robust model 2 for data set #2. For all results, the response is in logged units.

State/species/tree type	Confidence interval for bootstrap		Classical p-value
	99%	95%	
Georgia/slash/PSG1	(-0.3625, -0.0467)	(-0.3143, -0.0778)	0.0002
Georgia/slash/PSG5	(-0.3426, -0.0278)	(-0.2933, -0.0535)	0.0010
Alabama/shortleaf/PSG1	(-0.3733, -0.0280)	(-0.3511, -0.0658)	0.0030
Alabama/shortleaf/PSG5	(-0.2990, 0.0216)	(-0.2569, -0.0051)	0.0394
Georgia/shortleaf/PSG1	(-0.5428, -0.1798)	(-0.4930, -0.2313)	0.0000
Georgia/shortleaf/PSG5	(-0.4637, -0.1180)	(-0.4164, -0.1440)	0.0000
Alabama/loblolly/PSG1	(-0.2336, 0.0333)	(-0.2210, -0.0128)	0.0160
Alabama/loblolly/PSG5	(-0.1465, 0.0840)	(-0.1170, 0.0698)	0.5578
Georgia/loblolly/PSG1	(-0.3376, -0.1564)	(-0.3194, -0.1705)	0.0000
Georgia/loblolly/PSG5	(-0.2836, -0.1124)	(-0.2675, -0.1320)	0.0000
Alabama/longleaf/PSG1	(-0.5565, -0.1664)	(-0.5330, -0.2030)	0.0000
Alabama/longleaf/PSG5	(-0.3863, -0.0285)	(-0.3467, -0.0694)	0.0033
Georgia/longleaf/PSG1	(-0.3526, 0.0399)	(-0.2833, 0.0125)	0.0507
Georgia/longleaf/PSG5	(-0.2817, 0.0414)	(-0.2359, 0.0013)	0.0359

Table 9. Confidence intervals of $(b_1 - b_0)$ for bootstrap samples (1000 samples) and classical p-values for testing $H_0: (b_1 - b_0) = 0$ for the gross growth robust model 2a for data set #2. For all results, the response is in logged units.

State/species/tree type	Confidence interval for bootstrap		Classical p-value
	99%	95%	
Georgia/slash/PSG1	(-0.3897, -0.0690)	(-0.3416, -0.1039)	0.0001
Georgia/slash/PSG5	(-0.3578, -0.0456)	(-0.3099, -0.0690)	0.0005
Alabama/shortleaf/PSG1	(-0.5299, -0.1230)	(-0.4936, -0.1783)	0.0000
Alabama/shortleaf/PSG5	(-0.4313, -0.0772)	(-0.4125, -0.1140)	0.0002
Georgia/shortleaf/PSG1	(-0.5986, -0.2072)	(-0.5459, -0.2483)	0.0000
Georgia/shortleaf/PSG5	(-0.5098, -0.1486)	(-0.4596, -0.1838)	0.0000
Alabama/loblolly/PSG1	(-0.3045, -0.0266)	(-0.2899, -0.0515)	0.0006
Alabama/loblolly/PSG5	(-0.1975, 0.0446)	(-0.1714, 0.0313)	0.0907
Georgia/loblolly/PSG1	(-0.3710, -0.1485)	(-0.3480, -0.1826)	0.0000
Georgia/loblolly/PSG5	(-0.3171, -0.1227)	(-0.3017, -0.1467)	0.0000
Alabama/longleaf/PSG1	(-0.5686, -0.1840)	(-0.5381, -0.2117)	0.0000
Alabama/longleaf/PSG5	(-0.3968, -0.0382)	(-0.3500, -0.0809)	0.0018
Georgia/longleaf/PSG1	(-0.3393, 0.0620)	(-0.2672, 0.0362)	0.1267
Georgia/longleaf/PSG5	(-0.2764, 0.0416)	(-0.2229, 0.0095)	0.0686

Table 10. *Estimated gross growth from the gross growth robust model 2 for data set #2. All results are in original units.*

State/species/tree type	Estimated net growth 1st period	Estimated net growth 2nd period	Decrease (%)
Georgia/slash/PSG1	1.4034	1.0791	23.11
Georgia/slash/PSG5	1.0765	0.8588	20.22
Alabama/shortleaf/PSG1	2.8392	2.5937	8.65
Alabama/shortleaf/PSG5	2.3595	1.8107	23.26
Georgia/shortleaf/PSG1	1.0740	0.8458	21.25
Georgia/shortleaf/PSG5	1.1125	0.8163	26.62
Alabama/loblolly/PSG1	3.2718	3.2005	2.18
Alabama/loblolly/PSG5	1.2037	1.1682	2.95
Georgia/loblolly/PSG1	3.7226	3.1649	14.98
Georgia/loblolly/PSG5	1.6529	1.2977	21.49
Alabama/longleaf/PSG1	1.5823	1.1129	29.67
Alabama/longleaf/PSG5	0.7757	0.6019	22.40
Georgia/longleaf/PSG1	2.8804	2.7959	2.93
Georgia/longleaf/PSG5	0.9855	0.8074	18.07

Table 11. *Estimated gross growth from the gross growth robust model 2a for data set #2. All results are in original units.*

State/species/tree type	Estimated net growth 1st period	Estimated net growth 2nd period	Decrease (%)
Georgia/slash/PSG1	1.2856	0.8858	31.10
Georgia/slash/PSG5	1.0676	0.8442	20.93
Alabama/shortleaf/PSG1	2.9656	2.4353	17.88
Alabama/shortleaf/PSG5	2.4475	1.7844	27.09
Georgia/shortleaf/PSG1	1.0814	0.7318	32.33
Georgia/shortleaf/PSG5	1.0810	0.7026	35.00
Alabama/loblolly/PSG1	3.3315	3.1204	6.34
Alabama/loblolly/PSG5	1.1170	0.9861	11.72
Georgia/loblolly/PSG1	3.8290	3.1708	17.19
Georgia/loblolly/PSG5	1.6262	1.1964	26.43
Alabama/longleaf/PSG1	1.5803	1.0804	31.63
Alabama/longleaf/PSG5	0.7683	0.5772	24.87
Georgia/longleaf/PSG1	2.8581	2.7864	2.51
Georgia/longleaf/PSG5	0.9359	0.7870	15.91

bama-loblolly-PSG5, Georgia-longleaf-PSG1, and Georgia-longleaf-PSG5 for gross growth model 2; the bootstrap test and classical result showed significant differences except for Alabama-loblolly-PSG5, Georgia-longleaf-PSG1, and Georgia-longleaf-PSG5 for gross growth model 2a. Gross growth during the 2nd period ranged from 2% to 30% less than period one growth for model 2 and from 3% to 35% less for model 2a in Georgia-Alabama (tables 10 and 11).

Summary

This paper confirms the results of Bechtold *et al.* and Ruark *et al.* in addition to finding a decrease in growth rates in net growth for the Bechtold *et al.* data sets. The same models used by Bechtold *et al.* and Ruark *et al.* were adopted to analyze gross growth. A generalized linear modeling procedure was used to determine a model for net growth. Although a "true" model is hard to find, we believe the GLM approach is a sensible technique to determine the form of the model for net growth.

The procedure we describe in this paper is just one of many robust procedures. There is no one best procedure to use for all situations, and we believe the procedure used here is reasonable for the data analyzed. The iterative nature of the procedure allows detection of outliers for various models until the algorithm converges and a single "best" model is computed. Although outlying observa-

tions are modified at each iteration, data are not actually deleted as is the case for some techniques dealing with outliers. This robust procedure is sensible for data sets 1 and 2 because some data exploration indicated the possible existence of outlying observations.

We tested for a drop in tree growth from the 1st to the 2nd periods using a nonparametric bootstrapping technique that does not require distributional assumptions. We also used a classical test in order to compare the differences between a standard parametric test and the nonparametric test. We observed no significant differences between the two. The tests produced results that are inferences to unknown population parameters, which leads us to one last comment regarding the extent to which the data represents the population.

Although the criteria used to select plots were the same for both periods, we have no proof that the samples taken in the two periods represent the same population and we do not know the populations of inference. Actual populations of inference to which these results apply are unknown because sample plots were screened carefully to make the samples for the two periods comparable.

References

Bechtold, W. A., Ruark, G. A., and Lloyd F. T. 1991. Analyses of basal area growth reductions in Georgia's natural pine stands. Forest Science. 37: 703-732.

Chiu, S-T 1995. Personal communication. Colorado State University Department of Statistics.

Hastie, T. J. and Tibshirani, R. J. 1990. Generalized additive models. Chapman and Hall, New York.

Nelder, J. A. and Wedderburn, R.W.M. 1972. Generalized linear models. Journal of the Royal Statistical Society. Series A. 135: 370-384.

Ouyang, Z., Schreuder, H. T., and Li, H. G. 1992. A reevaluation of the growth decline in Georgia and Georgia-Alabama. Proc. Conf. on Applied Statistics in Agriculture April 28-30, 1991, Kansas State Univ. Manhattan, Kansas, p. 54-61

Pregibon, D. 1980. Goodness of link tests for generalized linear models. Applied Statistics. 29: 15-24.

Ruark, G. A., Thomas, C. E., Bechtold, W. A., and May, D. B. 1991. Growth reductions in naturally regenerated southern pine stands in Alabama and Georgia. Southern Journal of Applied Forestry. 15: 73-79.

Sheffield, R. M. and Cost, N. D. 1987. Behind the decline. Journal of Forestry. 85: 29-33.

Sheffield, R. M., Cost, N. D., Bechtold, W. A., and McClure, J. P. 1985. Pine growth reductions in the Southeast. USDA Forest Service Resource Bulletin SE 83. 112 pp.

Van Deusen, P. C. 1992. Growth trends and stand dynamics in natural loblolly pine in the southeastern United States. Canadian Journal of Forest Research. 22: 660-666.

Printed on ♻ recycled paper

Rocky Mountains

Southwest

Great Plains

U.S. Department of Agriculture
Forest Service

Rocky Mountain Forest and Range Experiment Station

The Rocky Mountain Station is one of seven regional experiment stations, plus the Forest Products Laboratory and the Washington Office Staff, that make up the Forest Service research organization.

RESEARCH FOCUS

Research programs at the Rocky Mountain Station are coordinated with area universities and with other institutions. Many studies are conducted on a cooperative basis to accelerate solutions to problems involving range, water, wildlife and fish habitat, human and community development, timber, recreation, protection, and multiresource evaluation.

RESEARCH LOCATIONS

Research Work Units of the Rocky Mountain Station are operated in cooperation with universities in the following cities:

Albuquerque, New Mexico
Flagstaff, Arizona
Fort Collins, Colorado*
Laramie, Wyoming
Lincoln, Nebraska
Rapid City, South Dakota

*Station Headquarters: 240 W. Prospect Rd., Fort Collins, CO 80526